



1. Rapid, accurate propagation of annotations
2. What should be the goal of our annotation efforts?
3. Subsystems Annotations

by

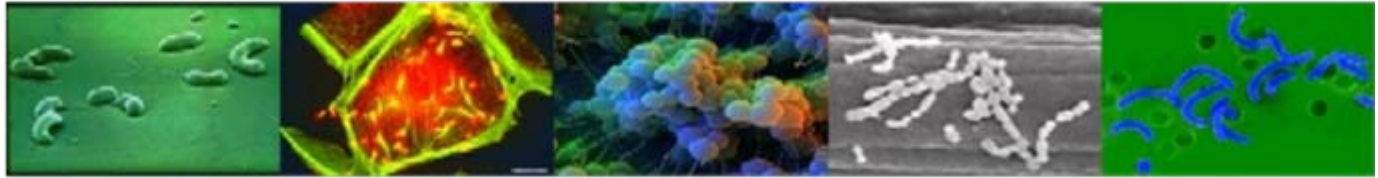
Ross Overbeek



NMPDR



National Microbial  
Pathogen Data  
Resource Center



# Rapid Propagation of annotations



THE UNIVERSITY OF  
CHICAGO



FELLOWSHIP FOR  
INTERPRETATION OF  
GENOMES



National Institute of Allergy and Infectious Diseases  
National Institutes of Health



NCSA



ARGONNE  
NATIONAL LABORATORY

# Basic Concepts for Annotating Closely-Related Strains

- Annotate one or more strains carefully
- Construct a set of “families”
- Make the families consistent
- Develop tools for rapidly propagating the family annotations to a new genome
- Identify portions not covered (and defer these for further analysis)



# What Must Be in the Families?

- Each family is a set of orthologs from existing, well-annotated genomes
- Each family has an associated function that applies to all members



# Required Tools

- A tool to produce initial families (both CDSs and RNAs)
- A tool to propagate families to a newly-sequenced genome
- A tool to determine genes not identified by propagation of families, with or without associated function assignments (these genes must be clearly separated from those produced by propagation)



# What is Achievable?

- We can, for many of our prokaryotic pathogens, produce accurate annotations of approximately 90% of the genes within 1-2 days.
- These annotations then become the starting point of detailed manual curation
- This becomes a key component in projects to sequence and annotate hundreds of closely-related strains





# What Are the Goals of Carefully Done Manual Curation?

1. Correct assignment of function
2. Connect to literature
3. Establish consistent protein families
4. Develop a metabolic reconstruction
5. Specify GO terms



# So, how do we reach these goals?

- UniProt curates protein families in which the members all have identical domain structures
- These families are decomposed into subfamilies (when it can be done) to separate distinct functions
- We should support construction of subfamilies with proteins that implement identical functions and have appropriate GO terms attached



# Why?

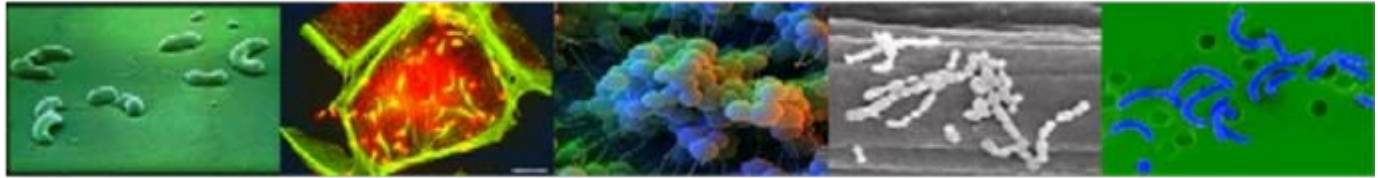
- It produces consistent annotations
- It minimizes the cost of attaching accurate GO terms
  - We can take the position that UniProt has the responsibility, or
  - We can actively provide the GO terms, and the impact would go beyond our specific organisms
  - Michael Ashburner from GO expressed interest in linking subsystem roles and GO terms
- It dramatically increases the impact of our manual curation efforts



NMPDR



National Microbial  
Pathogen Data  
Resource Center



# Subsystem Annotation: Why it is important



THE UNIVERSITY OF  
CHICAGO



FELLOWSHIP FOR  
INTERPRETATION OF  
GENOMES



National Institute of Allergy and Infectious Diseases  
National Institutes of Health



NCSA



ARGONNE  
NATIONAL LABORATORY

# What is a Subsystem?

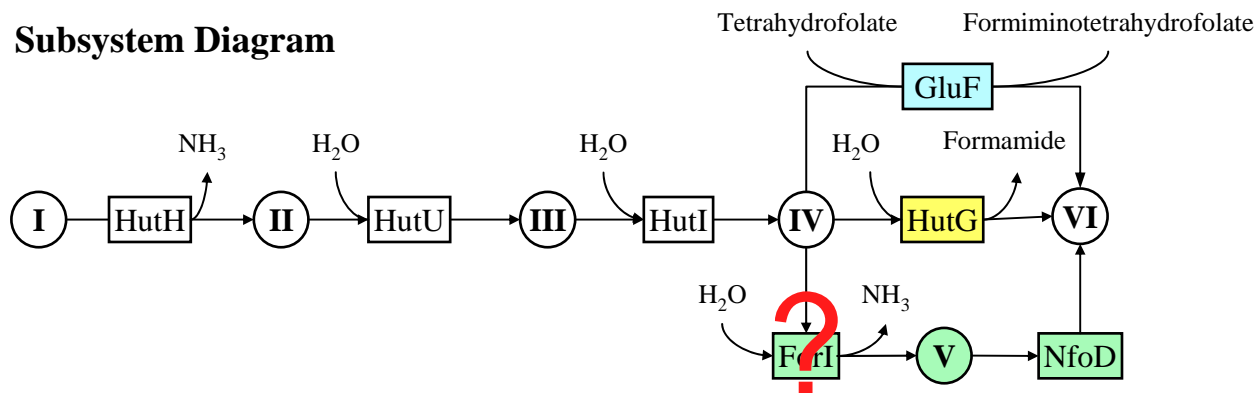
## Subsystem: Histidine Degradation

1	<b>HutH</b>	Histidine ammonia-lyase (EC 4.3.1.3)
2	<b>HutU</b>	Urocanate hydratase (EC 4.2.1.49)
3	<b>HutI</b>	Imidazolonepropionase (EC 3.5.2.7)
4	<b>GluF</b>	Glutamate formiminotransferase (EC 2.1.2.5)
5	<b>HutG</b>	Formiminoglutamase (EC 3.5.3.8)
6	<b>NfoD</b>	N-formylglutamate deformylase (EC 3.5.1.68)
7	<b>ForI</b>	Formiminoglutamic iminohydrolase (EC 3.5.3.13)

## Subsystem Spreadsheet

Organism	Variant	HutH	HutU	HutI	GluF	HutG	NfoD	ForI
<i>Bacteroides thetaiotaomicron</i>	1	<a href="#">Q8A4B3</a>	<a href="#">Q8A4A9</a>	<a href="#">Q8A4B1</a>	<a href="#">Q8A4B0</a>			
<i>Desulfotela psychrophila</i>	1	<a href="#">gi51246205</a>	<a href="#">gi51246204</a>	<a href="#">gi51246203</a>	<a href="#">gi51246202</a>			
<i>Halobacterium sp.</i>	2	<a href="#">Q9HQD5</a>	<a href="#">Q9HQD8</a>	<a href="#">Q9HQD6</a>		<a href="#">Q9HQD7</a>		
<i>Deinococcus radiodurans</i>	2	<a href="#">Q9RZ06</a>	<a href="#">Q9RZ02</a>	<a href="#">Q9RZ05</a>		<a href="#">Q9RZ04</a>		
<i>Bacillus subtilis</i>	2	<a href="#">P10944</a>	<a href="#">P25503</a>	<a href="#">P42084</a>		<a href="#">P42068</a>		
<i>Caulobacter crescentus</i>	3	<a href="#">P58082</a>	<a href="#">Q9A9MI</a>	<a href="#">P58079</a>			<a href="#">Q9A9M0</a>	<a href="#">Q9A9L9</a>
<i>Pseudomonas putida</i>	3	<a href="#">Q88CZ7</a>	<a href="#">Q88CZ6</a>	<a href="#">Q88CZ9</a>			<a href="#">Q88D00</a>	<a href="#">Q88CZ3</a>
<i>Xanthomonas campestris</i>	3	<a href="#">Q8PAA7</a>	<a href="#">P58988</a>	<a href="#">Q8PAA6</a>			<a href="#">Q8PAA8</a>	<a href="#">Q8PAA5</a>
<i>Listeria monocytogenes</i>	-1							

## Subsystem Diagram



# Example 2:

## Leucine Degradation

- Clarity through integration of
  - Metabolic context
  - Chromosomal context
  - Targeted wet lab confirmations
  - Projection of results

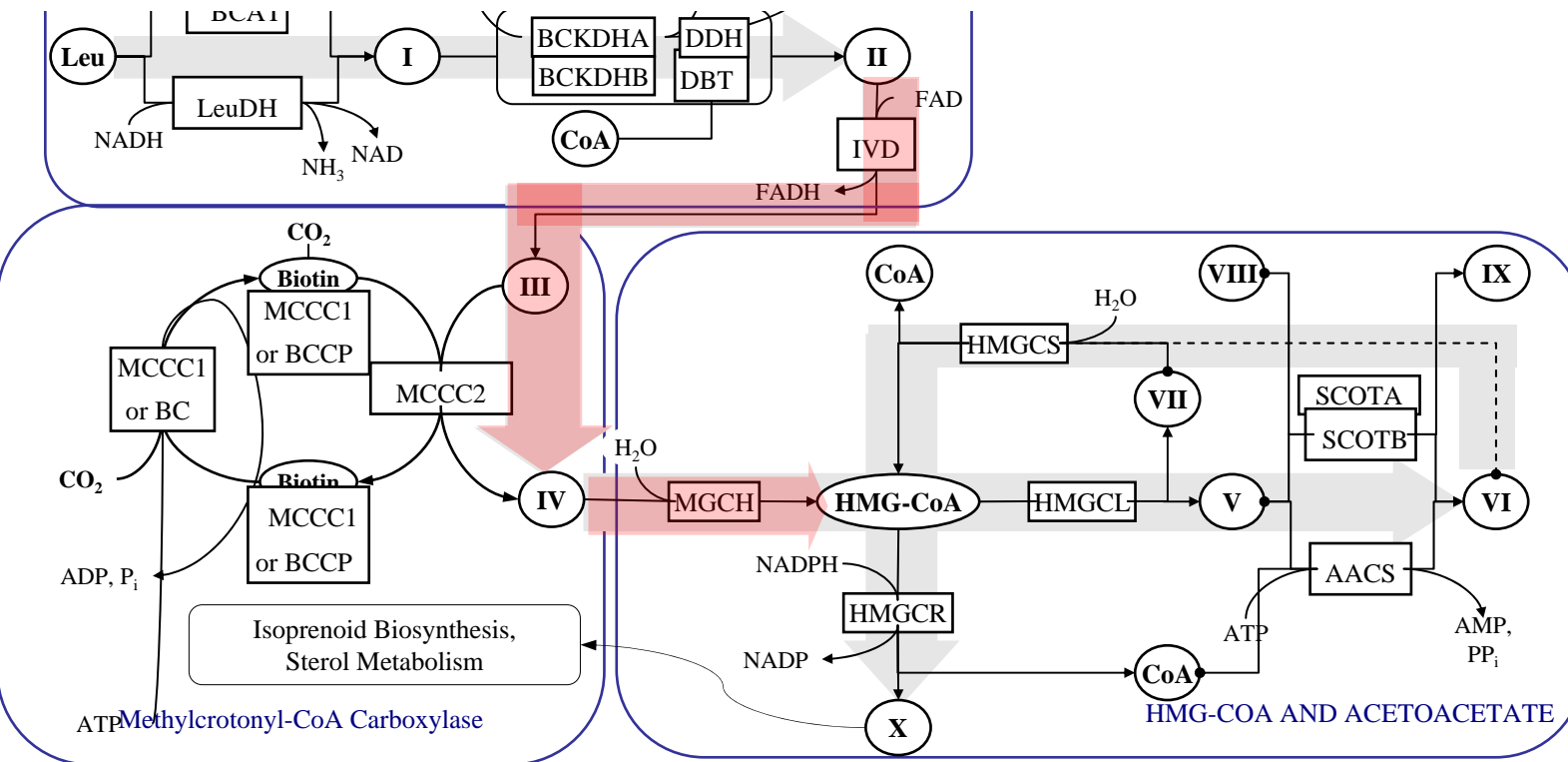


# Leucine Degradation and

Subsystem: Leucine Degradation and HMG-CoA Metabolism

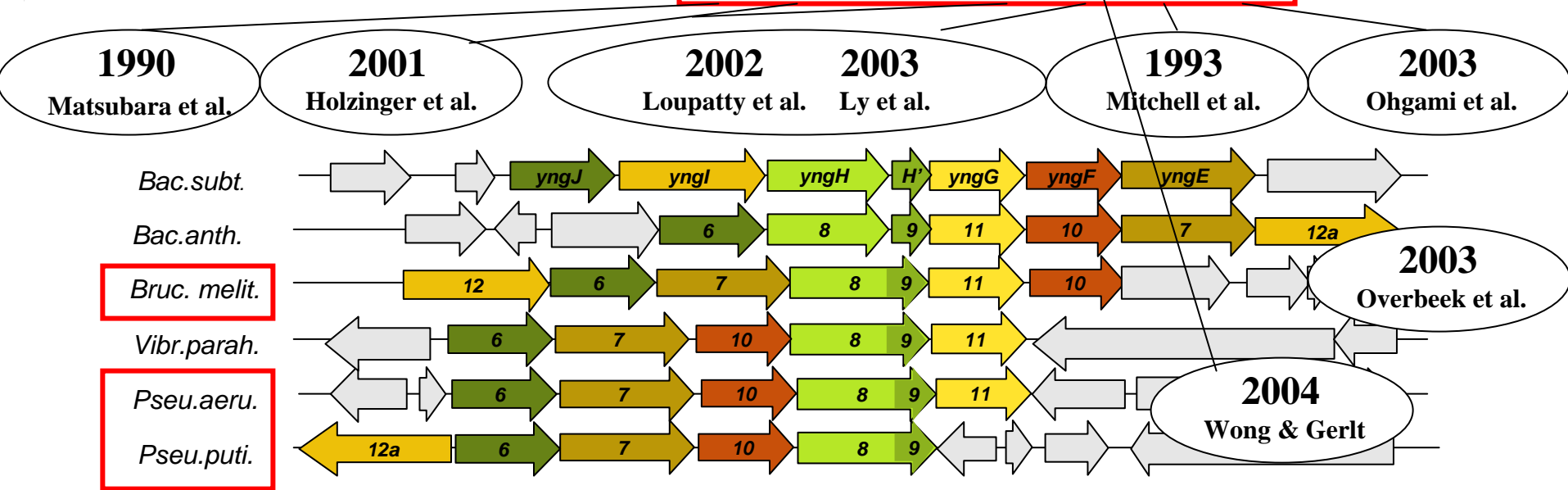
1 BCAT Branched-chain amino acid aminotransferase (EC 2.6.1.42)

7	<b>IVD</b>	Isovaleryl-CoA dehydrogenase (EC 1.3.99.10)
8	<b>MCCC1</b>	Methylcrotonyl-CoA carboxylase biotin-containing subunit (EC 6.4.1.4)
9	<b>BC</b>	Biotin carboxylase of methylcrotonyl-CoA carboxylase (EC 6.3.4.14)
10	<b>BCCP</b>	Biotin carboxyl carrier protein of methylcrotonyl-CoA carboxylase
11	<b>MCCC2</b>	Methylcrotonyl-CoA carboxylase carboxyl transferase subunit (EC 6.4.1.4)
12	<b>MGCH</b>	Methylglutaconyl-CoA hydratase (EC 4.2.1.18)
13	<b>HMGCL</b>	Hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4)
14	<b>AACS</b>	Acetoacetyl-CoA synthetase (EC 6.2.1.16)



# Leucine Degradation and HMG-CoA Metabolism

Genomes		From Leucine to Isovaleryl-CoA						To Acetoacetate via HMG-CoA							Acetoacetate		HMG-CoA		
		BCAT	LeuDH	BCKDHA	BCKDHB	DBT	DDH	IVD (6)	MCCC1 (8/9)	BC (8)	BCCP (9)	MCCC2 (7)	MGCH (10)	HMGCL (11)	AACS (12)	SCOTA (13)	SCOTB (14)	HMGCS	HMGCR
Bac.subtil.	1	ywaA	bcd	bkdA1	bkdA2	bkdB	yqiV	yngJ		yngH	yngH'	yngE	yngF	yngG	yngI	yxjD	yxjE	pksG	
Bac.anthr.	4	+	+	+	+	+	+	+		+	+	+	+	+	+				
Bruc.melit.	5	+		+	+	+	+	+	+			+	+	+	+				
Vibr.parah.	3	+		+	+	+	+	+	+			+	+	+	+	+	+		+
Pseu.aerug.	2	+	+	+	+	+	+	+	+			+	+	+	+	+	+		
Pseu.putida	2	+	+	+	+	+	+	+	+			+	+	+	+	+	+		
Hom.sapiens	6	+		+	+	+		+	+			+	+	+	+	+	+	+	+



## Leucine Degradation gene cluster in *Brucella suis* 1330

	<b>TIGR Annotation</b>	<b>TIGR Cellular role</b>	<b>Swiss-Prot / UniProt</b>	<b>GO</b>
Isovaleryl-CoA dehydrogenase (EC 1.3.99.10)	BR0020 isovaleryl-CoA dehydrogenase	Energy metabolism: Amino acids and amines	Isovaleryl-CoA dehydrogenase	GO:0006118; Biological process: electron transport GO:0016491; Molecular function: oxidoreductase activity GO:0008470; Molecular function: isovaleryl-CoA dehydrogenase activity
Methylcrotonyl-CoA carboxylase biotin-containing subunit (EC 6.4.1.4)	BR0018 biotin carboxylase	Unknown function: Enzymes of unknown specificity	Biotin carboxylase	GO:0005524; Molecular function: ATP binding GO:0016874; Molecular function: ligase activity GO:0008152; Biological process: metabolism
Methylcrotonyl-CoA carboxylase carboxyl transferase subunit (EC 6.4.1.4)	BR0019 carboxyl transferase family protein	Unknown function: Enzymes of unknown specificity	Carboxyl transferase family protein	GO:0016874; Molecular function: ligase activity GO:0016740; Molecular function: transferase activity
Methylglutaconyl-CoA hydratase (EC 4.2.1.18)	BR0016 enoyl- CoA hydratase/isomer ase family protein	Fatty acid and phospholipid metabolism: Degradation	Q8G3D2 Enoyl- CoA hydratase/isomer ase family protein	GO:0016853; Molecular function: isomerase activity GO:0008152; Biological process: metabolism
Hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4)	BR0017 3- hydroxy-3- methylglutarate- CoA lyase, authentic frameshift	Energy metabolism: Other	none	
Acetoacetyl-CoA synthetase (EC 6.2.1.16)	BR0021 acetoacetyl-CoA synthase	Central intermediary metabolism: Other	Acetoacetyl-CoA synthase	GO:0008152; Biological process: metabolism GO:0008299; Biological process: isoprenoid biosynthesis GO:0016874; Molecular function: ligase activity GO:0016405; Molecular function: CoA-ligase activity GO:0030729; Molecular function: acetoacetate-CoA ligase



# Participating Communities

- Those seeking better annotations as an end in themselves
- Those seeking “missing genes”
- The minimal organism community
- The pathogen community
- Those pursuing issues in phylogeny
- Those analyzing environmental samples
- Those building whole genome metabolic models



# How much progress has been made?

- 80 – 85% of the genes in core machinery are contained in subsystems
- 30 – 35% of genes in NMPDR organism genomes and 20 – 30% of other genomes contained in subsystems



## Virulence-related Subsystems in NMPDR

- **Attachement and colonization factors;**
    - *Pili and fimbriae*
    - *Nonfimbriar adhesins*
  - **Multiplication and Nutrition in host:**
    - *Iron scavenging mechanisms*
  - **Invasion and intracellular survival**
  - **Evading host defense mechanisms**
    - *Capsules, extracellular polymers*
    - *Antigenic variation*
  - **Antibiotic resistance**
  - **Toxins**
  - **Pathogen. islands, prophages, gene clusters**
  - **Other virulence factors:**
    - *Regulation of virulence*
    - *Secretion systems*
- Collaborations with research community have started
- **En**

	<i>Staphylo-ccocus</i>	<i>Strepto-ccocus</i>	<i>Listeria</i>	<i>Campylo-bacter</i>	<i>Vibrio</i>
		●	●	●	●
	●	●	●		
		●	●		
		●			●
		●	●		
●●●●●	●●●●●	●●	●●	●●	●
	●●			●	●
	●		●		●
● ●  ●	●●●●● ●  ●	● ●  ●	● ●  ●	●●  ● ●	●   ●
	●				

# Summary

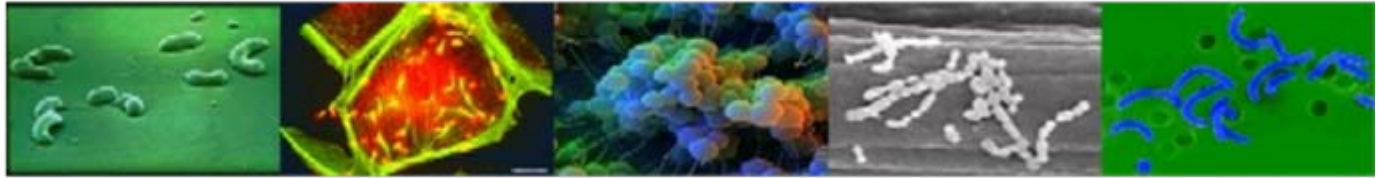
1. We can build a framework for very rapid, reasonably accurate annotation of genomes that are closely-related strains of those we carefully annotate. We should do so.
2. Our overall strategy of annotation and propagation of annotations should make use of the UniProt effort to construct families of genes with a common domain structure. We should seek subfamilies with a common cellular function and well-developed GO terms attached
3. Subsystems annotation will be the key technology used to remove many of the remaining ambiguities and errors



NMPDR



National Microbial  
Pathogen Data  
Resource Center



# What Follows are just some examples of the use of subsystems



THE UNIVERSITY OF  
CHICAGO



FELLOWSHIP FOR  
INTERPRETATION OF  
GENOMES



National Institute of Allergy and Infectious Diseases  
National Institutes of Health



NCSA



ARGONNE  
NATIONAL LABORATORY

# Rubisco in *Bacillus anthracis*?

[Show/Hide Assignments for Essentially Identical Proteins](#)

Assignments for Essentially Identical Proteins

Id	Organism	Who	ASSIGN	Assignment
<a href="#">figl260799.1.peg.3905</a>	Bacillus anthracis str. Sterne			2,3-diketo-5-methylthiopentyl-1-phosphate enolase (EC 5.3.2.-)
<a href="#">figl198094.1.peg.3918</a>	Bacillus anthracis str. Ames	FIG		2,3-diketo-5-methylthiopentyl-1-phosphate enolase (EC 5.3.2.-)
<a href="#">figl261594.1.peg.4224</a>	Bacillus anthracis str. 'Ames Ancestor	FIG		2,3-diketo-5-methylthiopentyl-1-phosphate enolase (EC 5.3.2.-)
<a href="#">gil65321428</a>	Bacillus anthracis str. A2012		<=	COG1850: Ribulose 1,5-bisphosphate carboxylase, large subunit
<a href="#">keggbaa:BA_4714</a>	Bacillus anthracis A2012	KEGG	<=	ribulose bisphosphate carboxylase large chain, catalytic domain [EC:4.1
<a href="#">keggban:BA4255</a>	Bacillus anthracis Ames	KEGG	<=	ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39] [KO:K0160
<a href="#">keggbar:GBAA4255</a>	Bacillus anthracis Ames 0581	KEGG	<=	ribulose bisphosphate carboxylase, putative [EC:4.1.1.39] [KO:K01601
<a href="#">keggbat:BAS3946</a>	Bacillus anthracis Sterne	KEGG	<=	ribulose bisphosphate carboxylase, putative [EC:4.1.1.39] [KO:K01601
<a href="#">tigrBA4255</a> (Pathema)	Bacillus anthracis Ames	TIGR	<=	ribulose bisphosphate carboxylase, putative
<a href="#">tigrGBAA4255</a> (Pathema)	Bacillus anthracis Ames Ancestor	TIGR	<=	ribulose bisphosphate carboxylase, putative

Context-  
based  
Homology-  
based

## I. Homology- based annotation (via protein families)

[To View Annotations](#) / [To View All Related Annotations](#)  
[Edit Controlled Vocabulary](#)  
[Protein Sequence](#)  
[DNA Sequence](#)

[To Compare Region](#)

[Explore Protein Families for figl260799.1.peg.3905](#)

Family	Family Function	External IDs In Family	Unique Proteins In Family
<a href="#">figlPF001312</a>	2,3-diketo-5-methylthiopentyl-1-phosphate enolase (EC 5.3.2.-)	22	20
<a href="#">pfamPF00016.9</a>	RuBisCO_large	14950	13297
<a href="#">spIPS00157</a>	RUBISCO_LARGE	545	515

## II. Genome context- and functional context- based annotation (via subsystems)

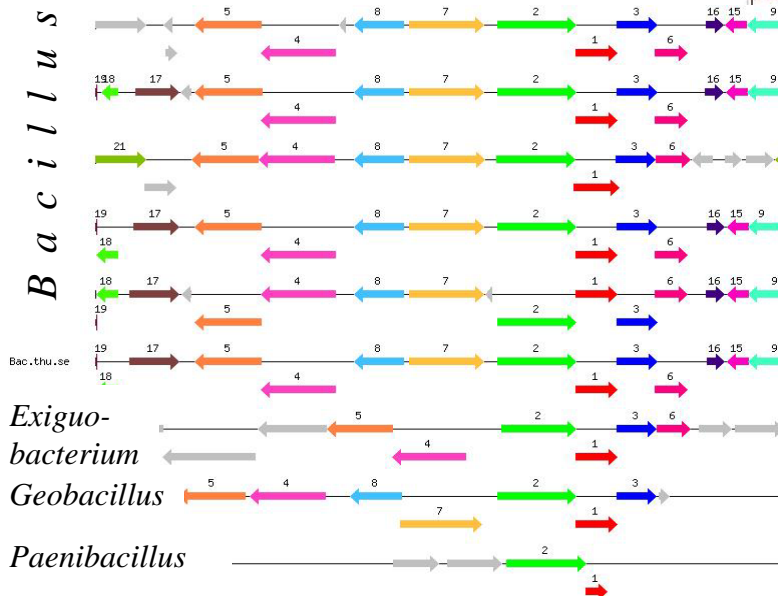
### Subsystem: Calvin-Benson cycle

Genome ID	Organism	Variant Code	RbcL	RbcS	PKK	*GAPDH	TPI	*FBA	*FBP	SBP	*TK	RPE	*Ris	PRK	CbbR	RPDH
198094.1 <a href="#">Color</a>	Bacillus anthracis str. Ames [B]	-1			<a href="#">4969</a>	<a href="#">4458-4, 4970-5</a>	<a href="#">4968</a>	<a href="#">5164-9</a>	<a href="#">4806-11, 5161-11</a>		<a href="#">3144-14, 3433-14</a>	<a href="#">3668</a>	<a href="#">2568-18, 5146-19</a>			

Context on contig 260799.1:NC\_005945 from base 3887717 to 3898900 (11184)

fid	starts	ends	size	gap	find best clusters	pins	fc-sc	SS	Ev	comment		
<a href="#">3901</a>	3889828	3888773	1056	-			6	1	icw(1)	Methylthioribose-1-phosphate isomerase (EC 5.3.1.23)		
<a href="#">3902</a>	38891006	3889825	1182	-4			6	1	isu icw(2)	5-methylthioribose kinase (EC 2.7.1.100)		
<a href="#">3903</a>	3882090	3891311	780	-304			5			hydrolase, carbon-nitrogen family		
<a href="#">3904</a>	3882188	3893366	1179	+97			5	1	isu	Glutamine-dependent 2-keto-4-methylthiobutyrate transaminase (EC 2.6.1.-)		
<a href="#">3905</a>	3883603	3894847	1245	+236				1	isu icw(1)	2,3-diketo-5-methylthiopentyl-1-phosphate enolase (EC 5.3.2.-)		
<a href="#">3906</a>	3884844	3895503	560	+4			1	1	isu icw(1)	2-hydroxy-3-keto-5-methylthiopentenyl-1-phosphate phosphatase (EC 3.1.3.-)		
<a href="#">3907</a>	3885500	3896138	539	+4			6	1	isu icw(2)	Methylthioribulose-1-phosphate dehydratase (EC 4.2.1.-)		
<a href="#">3896116</a>	3896628	513	+23				6	1	isu icw(1)	1,2-dihydroxy-3-keto-5-methylthiopentono dioxygenase (EC 1.13.-.-)		
<a href="#">3896930</a>	3897210	276	+306									

### Rubisco homolog clusters with Methionine salvage - related genes:



### Subsystems in which this peg is present

Subsystem	Curator	Role
Methionine Salvage	gjo	2,3-diketo-5-methylthiopentyl-1-phosphate enolase (EC 5.3.2.-)

Agrees with experimental data:

[Sekowska A, Danchin A.](#)

The methionine salvage pathway in *Bacillus subtilis*.  
BMC Microbiol. 2002 Apr 25;2:8.  
PMID: 12022921 [PubMed - indexed for MEDLINE]

## Na(+) translocating oxidoreductase and rnf electron transport complexes

<b>FIG Function</b>	<b>Other Function</b>
Na(+)-translocating NADH-quinone reductase subunit E (EC 1.6.5.-)	hypothetical protein
Na(+)-translocating NADH-quinone reductase subunit E (EC 1.6.5.-)	putative oxidoreductase
Electron transport complex protein mfa	conserved hypothetical protein
Electron transport complex protein mfa	conserved hypothetical protein
Electron transport complex protein mfa	hypothetical protein
Electron transport complex protein mfa	hypothetical protein
Electron transport complex protein mfa	hypothetical protein
Electron transport complex protein mfa	putative membrane protein
Electron transport complex protein mfa	conserved hypothetical protein
Electron transport complex protein mfa	conserved hypothetical protein
Electron transport complex protein mfa	hypothetical protein
Electron transport complex protein mfa	hypothetical protein
Electron transport complex protein mfa	putative membrane protein
Electron transport complex protein mfa	putative membrane protein
Electron transport complex protein mfa	putative membrane protein
Electron transport complex protein mfC	putative membrane protein
Electron transport complex protein mfC	putative membrane protein
Electron transport complex protein mfD	hypothetical protein
Electron transport complex protein mfD	
Electron transport complex protein mfD	
Electron transport complex protein mfD	
Electron transport complex protein mfD	
Electron transport complex protein mfE	putative oxidoreductase
Electron transport complex protein mfE	conserved hypothetical protein
Electron transport complex protein mfE	conserved hypothetical protein
Electron transport complex protein mfE	conserved hypothetical protein
Electron transport complex protein mfE	hypothetical protein
Electron transport complex protein mfE	hypothetical protein
Electron transport complex protein mfE	orf conserved hypothetical protein
Electron transport complex protein mfF	conserved hypothetical protein

## Discrepancies in annotation of rnf -related-electron transport complexes

Annotated as  
hypothetical  
protein

## Other

# NMPDR

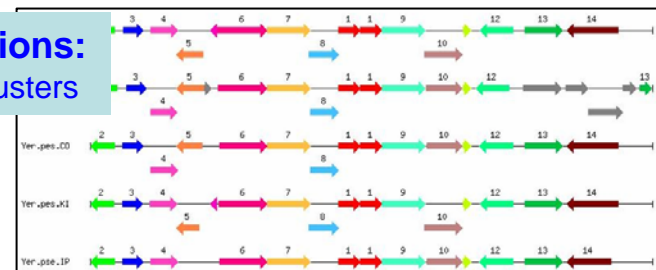
Systematic  
annotation of  
complex subunits

**Populated SS:** two types of complexes are distinguished by gene clustering

## Subsystem:

Column	Abbreu	Functional Role
1	A	Na(+)-translocating NADH-quinone reductase subunit A (EC 1.6.5.-)
2	B	Na(+)-translocating NADH-quinone reductase subunit B (EC 1.6.5.-)
3	C	Na(+)-translocating NADH-quinone reductase subunit C (EC 1.6.5.-)
4	D	Na(+)-translocating NADH-quinone reductase subunit D (EC 1.6.5.-)
5	E	Na(+)-translocating NADH-quinone reductase subunit E (EC 1.6.5.-)
6	F	Na(+)-translocating NADH-quinone reductase subunit F (EC 1.6.5.-)
7	D/E	Na(+)-translocating NADH-quinone reductase subunit D/E (EC 1.6.5.-)
8	mfA	Electron transport complex protein mfA
9	mfB	Electron transport complex protein mfB
10	mfC	Electron transport complex protein mfC
11	mfD	Electron transport complex protein mfD
12	mfE	Electron transport complex protein mfE
13	mfG	Electron transport complex protein mfG

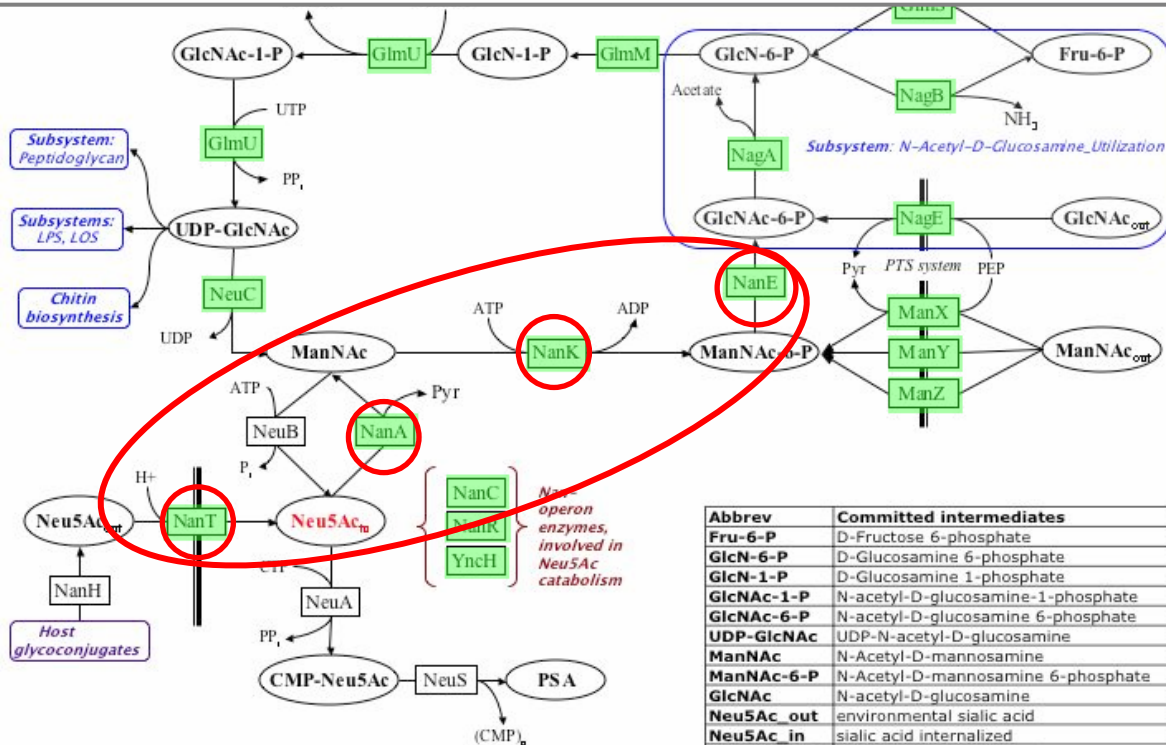
## Pinned regions: Conserved clusters



Genome ID	Organism	Variant	Code	A	B	C	*D	E	F	rnfA	rnfB	rnfC	rnfD	rnfE	rnfF
107806.1	Buchnera aphidicola str. APS (Acacia)									113	114	115	116	118	117
224915.1	Buchnera aphidicola str. Bp (Baobab)									98	99	100	101	103	102
198804.1	Buchnera aphidicola str. Sg (Schizanthus)									103	104	550	105	107	106
216592.1	Escherichia coli 042 [B]									2232	2233	2234	2235	2237	2236
199310.1	Escherichia coli CFT073 [B]									1957	1958	1959	1960	1962	1961
171440.1	Photorhabdus asymbiotica subsp. asymbiotica [B]	1		2433	593	2432	592-7	591	590	893	3008	892	891	889	890
630.2	Yersinia enterocolitica 8081 [B]	1		3149	3148	3147	3146-7	3145	3144	1943	1942	1941	1939	1937	1938
214092.1	Yersinia pestis CO92 [B]	1		3228	3227	3226	3225-7	2010, 3224	3223	2303	2302	2301	2299	2297	2298
198215.1	Shigella flexneri 2a str. 2457T [B]	7								1571	1572	1573	1574	1576	1575
198214.1	Shigella flexneri 2a str. 301 [B]	7								1567	1568	1569	1570	1572	1571

## ***Nan*-operon, a key pathway within the Sialic Acid Metabolism subsystem**

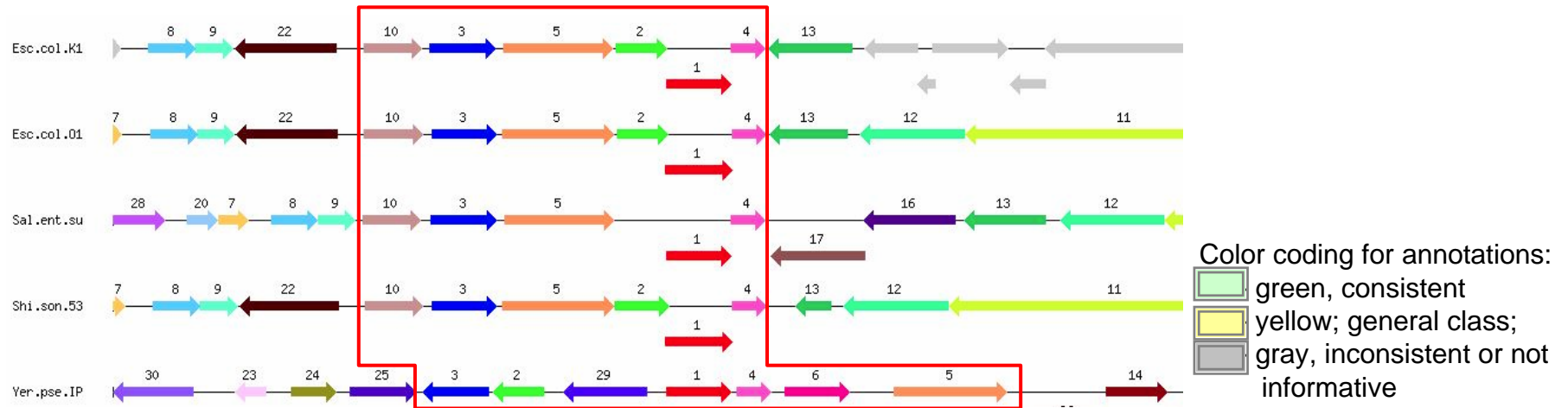
Genome ID	Organism	Variant Code	NanR	NanC	NanT	NanA	NanK	NanE	YhcH
83333.1	<a href="#">Color</a> Escherichia coli K12 [B]	2A	<a href="#">3170</a>	<a href="#">4221</a>	<a href="#">3168</a> , <a href="#">4187</a>	<a href="#">3169</a>	<a href="#">3166</a>	<a href="#">3167</a>	<a href="#">3165</a>
155864.1	<a href="#">Color</a> Escherichia coli O157:H7 EDL933 [B]	12A	<a href="#">4102</a>	<a href="#">5235</a>	<a href="#">4100</a>	<a href="#">4101</a>	<a href="#">4098</a>	<a href="#">4099</a>	<a href="#">4097</a>
83334.1	<a href="#">Color</a> Escherichia coli O157:H7 [B]	2A	<a href="#">4073</a>	<a href="#">5250</a>	<a href="#">4071</a>	<a href="#">4072</a>	<a href="#">4069</a>	<a href="#">4070</a>	<a href="#">4068</a>
321314.4	<a href="#">Color</a> Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 [B]	2A	<a href="#">3892</a>	<a href="#">672</a>	<a href="#">3890</a>	<a href="#">3891</a>	<a href="#">3888</a>	<a href="#">670</a> , <a href="#">3889</a>	<a href="#">3887</a>
594.1	<a href="#">Color</a> Salmonella enterica subsp. enterica serovar Gallinarum [B]	2A	<a href="#">2891</a>	<a href="#">4103</a>	<a href="#">2893</a> , <a href="#">4102</a>	<a href="#">2892</a>	<a href="#">2895</a>	<a href="#">2894</a> , <a href="#">4106</a>	<a href="#">2896</a>
209261.1	<a href="#">Color</a> Salmonella enterica subsp. enterica serovar Typhi Ty2 [B]	2A	<a href="#">3042</a>	<a href="#">1677</a>	<a href="#">1676</a> , <a href="#">3040</a>	<a href="#">3041</a>	<a href="#">3039</a>	<a href="#">1679</a>	<a href="#">3038</a>
99287.1	<a href="#">Color</a> Salmonella typhimurium LT2 [B]	2A	<a href="#">3226</a>	<a href="#">1094</a>	<a href="#">1095</a> , <a href="#">3224</a>	<a href="#">3225</a>	<a href="#">3222</a>	<a href="#">1092</a> , <a href="#">3223</a>	<a href="#">3221</a>
198214.1	<a href="#">Color</a> Shigella flexneri 2a str. 301 [B]	2A	<a href="#">3059</a>	<a href="#">3964</a>	<a href="#">3057</a>	<a href="#">3058</a>	<a href="#">3055</a>	<a href="#">3056</a>	<a href="#">3054</a>
229193.1	<a href="#">Color</a> Yersinia pestis biovar Medievalis str. 91001 [B]	2A			<a href="#">2551</a>	<a href="#">2558</a>	<a href="#">2555</a>	<a href="#">2557</a>	<a href="#">2554</a>
273123.1	<a href="#">Color</a> Yersinia pseudotuberculosis IP 32953 [B]	2A			<a href="#">2827</a>	<a href="#">2833</a>	<a href="#">2830</a>	<a href="#">2832</a>	<a href="#">2829</a>



Microbial sialic acid metabolism has now been firmly established as a virulence determinant in a range of infectious diseases

Abbrev	Committed intermediates
<b>Fru-6-P</b>	D-Fructose 6-phosphate
<b>GlcN-6-P</b>	D-Glucosamine 6-phosphate
<b>GlcN-1-P</b>	D-Glucosamine 1-phosphate
<b>GlcNAc-1-P</b>	N-acetyl-D-glucosamine-1-phosphate
<b>GlcNAc-6-P</b>	N-acetyl-D-glucosamine 6-phosphate
<b>UDP-GlcNAc</b>	UDP-N-acetyl-D-glucosamine
<b>ManNAc</b>	N-Acetyl-D-mannosamine
<b>ManNAc-6-P</b>	N-Acetyl-D-mannosamine 6-phosphate
<b>GlcNAc</b>	N-acetyl-D-glucosamine
<b>Neu5Ac_out</b>	environmental sialic acid
<b>Neu5Ac_in</b>	sialic acid internalized
<b>CMP-Neu5Ac</b>	CMP-N-acetylneuraminate
<b>PEP</b>	phosphoenolpyruvate
<b>PSA</b>	Polysialic acid

# Comparison of annotations within the conserved cluster (*nan*-operon)



No in cluster	Abbr.	Functional role in subsystem	Escherichia coli O157:H7 EDL933		Salmonella enterica subsp. enterica serovar Typhi Ty2		Shigella sonnei 53G		Yersinia pseudotuberculosis IP	
1	NanK	<b>N-acetylmannosamine kinase (EC 2.7.1.60)</b>	ABH-0028250	putative NAGC-like transcriptional regulator	ABS-0084973	possible kinase	ADD-0003671	putative NAGC-like transcriptional regulator	ACZ-0002834	putative sugar kinase
2	NanE	<b>N-acetylmannosamine-6-phosphate 2-epimerase (EC 5.1.3.9)</b>	ABH-0028251	putative enzyme	ABS-0083505	conserved hypothetical protein	ADD-0003672	putative enzyme	ACZ-0002836	conserved hypothetical protein
3	NanA	<b>N-acetylneuraminate lyase (EC 4.1.3.3)</b>	ABH-0028253	N-acetyl-neuraminate lyase	ABS-0084976	N-acetyl-neuraminate lyase	ADD-0003674	N-acetyl-neuraminate lyase	ACZ-0002837	probable N-acetylneuraminate lyase
4	YhcH	<b>Putative sugar isomerase involved in processing of exogenous sialic acid*</b>	ABH-0028249	orf, hypothetical protein	ABS-0084972	conserved hypothetical protein	ADD-0003670	conserved hypothetical protein	ACZ-0002833	conserved hypothetical protein
5	NanT	<b>Sialic acid transporter (permease) NanT</b>	ABH-0028252	sialic acid transporter	ABS-0084975	putative sialic acid transporter	ADD-0003673	sialic acid transporter	ACZ-0002831	MFS family sialic acid transporter
10	NanR	<b>Transcriptional regulator NanR**</b>	ABH-0028254	putative FADA-type transcriptional regulator	ABS-0084977	putative GntR-family transcriptional regulator	ADD-0003675	putative FADA-type transcriptional regulator	NOT PRESENT (likely repalced by a clustered member of RpiR family)	

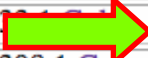
\* proposed by: 9. Teplyakov, A., Obmolova, G., Toedt, J., Galperin, M. Y., Gilliland, G. L. (2005). Crystal Structure of the Bacterial YhcH Protein Indicates a Role in Sialic Acid Catabolism. J. Bacteriol. 187: 5520-5527

\*\* K. A. Kalivoda, S. M. Steenbergen, E. R. Vimr, and J. Plumbridge  
Regulation of Sialic Acid Catabolism by the DNA Binding Protein NanR in Escherichia coli. J. Bacteriol., August 15, 2003; 185(16): 4806 - 4815

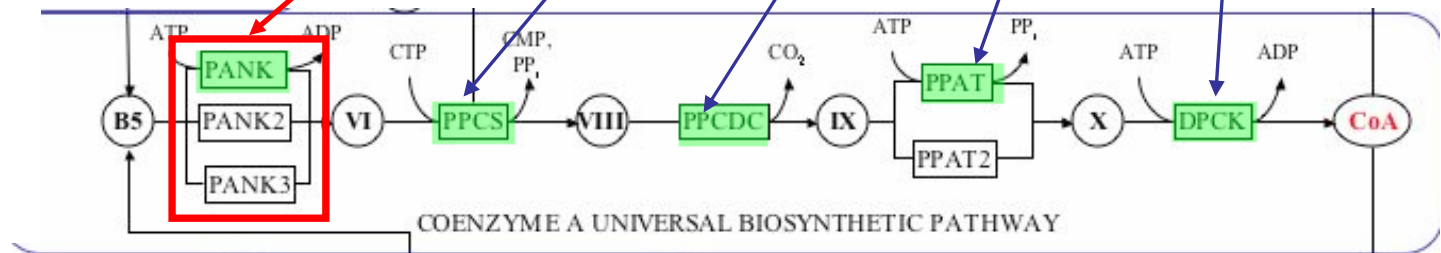
# Annotations of alternative pantothenate kinase genes in CoA\_Biosynthesis

Section of the subsystem capturing the “universal” pathway from pantothenate to CoA

7	PANK	Pantothenate kinase (EC 2.7.1.33)	<a href="#">R03018</a>
8	PANK2	Pantothenate kinase type II, eukaryotic (EC 2.7.1.33)	<a href="#">R03018</a>
9	PANK3	Pantothenate kinase type III, CoaX-like (EC 2.7.1.33)	<a href="#">R03018</a>
10	PPCS	Phosphopantothenoylcysteine synthetase (EC 6.3.2.5)	<a href="#">R04231</a>
11	PPCDC	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	<a href="#">R03269</a>
12	PPAT	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	<a href="#">R03035</a>
13	PPAT2	Phosphopantetheine adenylyltransferase, type II eukaryotic (EC 2.7.7.3)	<a href="#">R03035</a>
14	DPCCK	Dephospho-CoA kinase (EC 2.7.1.24)	<a href="#">R00130</a>

Genome ID	Organism	Variant Code	*PANK	PPCS	PPCDC	*PPAT	DPCCK	
833 	Escherichia coli K12 [B]	A	<b>3890-7</b>	<a href="#">3575</a>	<a href="#">3575</a>	<a href="#">3570-12</a>	<a href="#">103</a>	model organisms
224308.1 <a href="#">Color</a>	Bacillus subtilis subsp. subtilis str. 168 [B]	B	<a href="#">2381-7</a> , <a href="#">70-9</a>	<a href="#">1572</a>	<a href="#">1572</a>	<a href="#">1504-12</a>	<a href="#">2909</a>	
158879.1 <a href="#">Color</a>	Staphylococcus aureus subsp. aureus N315 [B]	C	<a href="#">1999-8</a>	<a href="#">1089</a>	<a href="#">1089</a>	<a href="#">1004-12</a>	<a href="#">1561</a>	
261594.1 <a href="#">Color</a>	Bacillus anthracis str. 'Ames Ancestor' [B]	C	<a href="#">2974-8</a> , <a href="#">370-9</a>	<a href="#">2999</a> , <a href="#">3983</a>	<a href="#">3983</a>	<a href="#">4112-12</a>	<a href="#">4765</a>	Pathema organisms
1491.1 <a href="#">Color</a>	Clostridium botulinum ATCC 3502 [B]	E	<a href="#">2254-9</a>	<a href="#">1240</a>	<a href="#">1240</a>	<a href="#">3223-12</a>	<a href="#">2054</a>	
272560.3 <a href="#">Color</a>	Burkholderia pseudomallei K96243 [B]	E	<a href="#">2992-9</a>	<a href="#">3248</a>	<a href="#">3248</a>	<a href="#">2660-12</a>	<a href="#">5587</a>	

PANK of CoA family, a canonical form characteristic for *E.coli* and many other bacteria



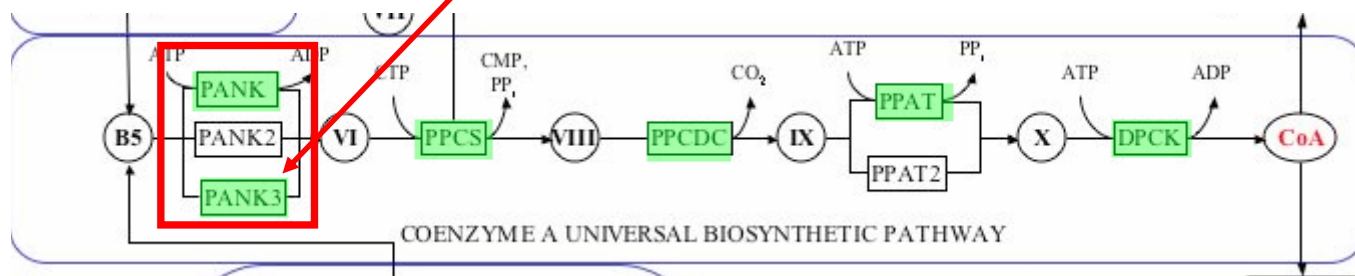
- The overall topology of this pathway is conserved in most species including archaea and eukaryotes;
- Major variations: alternative (nonhomologous) forms of Pantothenate Kinase (\*PANK)

# Annotations of alternative pantothenate kinase genes in CoA\_Biosynthesis

7	PANK	Pantothenate kinase (EC 2.7.1.33)	<a href="#">R03018</a>
8	PANK2	Pantothenate kinase type II, eukaryotic (EC 2.7.1.33)	<a href="#">R03018</a>
9	PANK3	Pantothenate kinase type III, CoaX-like (EC 2.7.1.33)	<a href="#">R03018</a>
10	PPCS	Phosphopantothenoylcysteine synthetase (EC 6.3.2.5)	<a href="#">R04231</a> ,
11	PPCDC	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	<a href="#">R03269</a>
12	PPAT	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	<a href="#">R03035</a>
13	PPAT2	Phosphopantetheine adenylyltransferase, type II eukaryotic (EC 2.7.7.3)	<a href="#">R03035</a>
14	DPCK	Dephospho-CoA kinase (EC 2.7.1.24)	<a href="#">R00130</a>

Genome ID	Organism	Variant Code	*PANK	PPCS	PPCDC	*PPAT	DPCK	
83333.1 <a href="#">Color</a>	Escherichia coli K12 [B]	A	<a href="#">3890-7</a>	<a href="#">3575</a>	<a href="#">3575</a>	<a href="#">3570-12</a>	<a href="#">103</a>	model organisms
224 <a href="#">Color</a>	Bacillus subtilis subsp. subtilis str. 168 [B]	B	<a href="#">2381-7</a> <a href="#">70-9</a>	<a href="#">1572</a>	<a href="#">1572</a>	<a href="#">1504-12</a>	<a href="#">2909</a>	
158879.1 <a href="#">Color</a>	Staphylococcus aureus subsp. aureus N315 [B]	C	<a href="#">1999-8</a>	<a href="#">1089</a>	<a href="#">1089</a>	<a href="#">1004-12</a>	<a href="#">1561</a>	
261594.1 <a href="#">Color</a>	Bacillus anthracis str. 'Ames Ancestor' [B]	C	<a href="#">2974-8</a> , <a href="#">370-9</a>	<a href="#">2999</a> , <a href="#">3983</a>	<a href="#">3983</a>	<a href="#">4112-12</a>	<a href="#">4765</a>	Pathema organisms
1491.1 <a href="#">Color</a>	Clostridium botulinum ATCC 3502 [B]	E	<a href="#">2254-9</a>	<a href="#">1240</a>	<a href="#">1240</a>	<a href="#">3223-12</a>	<a href="#">2054</a>	
272560.3 <a href="#">Color</a>	Burkholderia pseudomallei K96243 [B]	E	<a href="#">2992-9</a>	<a href="#">3248</a>	<a href="#">3248</a>	<a href="#">2660-12</a>	<a href="#">5587</a>	

An additional non-homologous PANK2 gene (CoaX) was identified in *B. subtilis* and other bacteria.



[J Biol Chem.](#) 2005 May 27;280(21):20185-8. Epub 2005 Mar 28.

Full text article at  
[www.jbc.org](http://www.jbc.org)

[Find It!](#)  
Burnham Access

Characterization of a new pantothenate kinase isoform

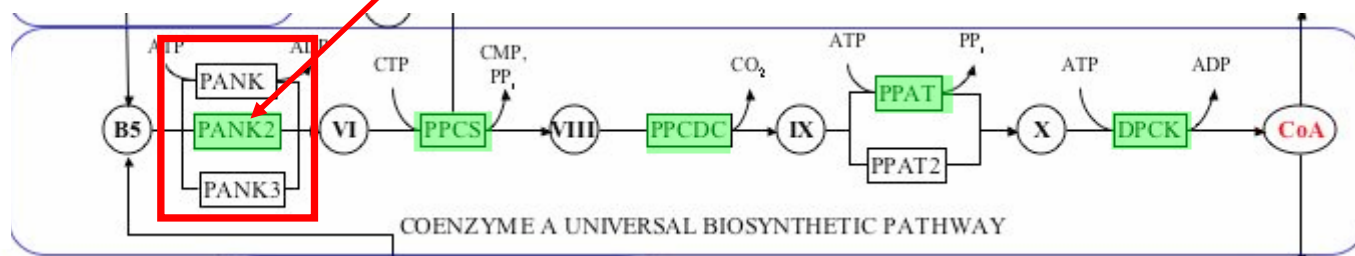
[Brand LA](#), [Strauss E](#).

# Annotations of alternative pantothenate kinase genes in CoA\_Biosynthesis

7	PANK	Pantothenate kinase (EC 2.7.1.33)	<a href="#">R03018</a>
8	PANK2	Pantothenate kinase type II, eukaryotic (EC 2.7.1.33)	<a href="#">R03018</a>
9	PANK3	Pantothenate kinase type III, CoaX-like (EC 2.7.1.33)	<a href="#">R03018</a>
10	PPCS	Phosphopantothenoylcysteine synthetase (EC 6.3.2.5)	<a href="#">R04231</a> ,
11	PPCDC	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	<a href="#">R03269</a>
12	PPAT	Phosphopantetheine adenyltransferase (EC 2.7.7.3)	<a href="#">R03035</a>
13	PPAT2	Phosphopantetheine adenyltransferase, type II eukaryotic (EC 2.7.7.3)	<a href="#">R03035</a>
14	DPCK	Dephospho-CoA kinase (EC 2.7.1.24)	<a href="#">R00130</a>

Genome ID	Organism	Variant Code	*PANK	PPCS	PPCDC	*PPAT	DPCK	
83333.1 <a href="#">Color</a>	Escherichia coli K12 [B]	A	<a href="#">3890-7</a>	<a href="#">3575</a>	<a href="#">3575</a>	<a href="#">3570-12</a>	<a href="#">103</a>	model organisms
224308.1 <a href="#">Color</a>	Bacillus subtilis subsp. subtilis str. 168 [B]	B	<a href="#">2381-7</a> , <a href="#">70-9</a>	<a href="#">1572</a>	<a href="#">1572</a>	<a href="#">1504-12</a>	<a href="#">2909</a>	
158 <a href="#">Color</a>	Staphylococcus aureus subsp. aureus N315 [B]	C	<a href="#">1999-8</a>	<a href="#">1089</a>	<a href="#">1089</a>	<a href="#">1004-12</a>	<a href="#">1561</a>	
261594.1 <a href="#">Color</a>	Bacillus anthracis str. 'Ames Ancestor' [B]	C	<a href="#">2974-8</a> , <a href="#">370-9</a>	<a href="#">2999</a> , <a href="#">3983</a>	<a href="#">3983</a>	<a href="#">4112-12</a>	<a href="#">4765</a>	Pathema organisms
1491.1 <a href="#">Color</a>	Clostridium botulinum ATCC 3502 [B]	E	<a href="#">2254-9</a>	<a href="#">1240</a>	<a href="#">1240</a>	<a href="#">3223-12</a>	<a href="#">2054</a>	
272560.3 <a href="#">Color</a>	Burkholderia pseudomallei K96243 [B]	E	<a href="#">2992-9</a>	<a href="#">3248</a>	<a href="#">3248</a>	<a href="#">2660-12</a>	<a href="#">5587</a>	

A distant homolog of a structurally unrelated eukaryotic PANK3 gene was inferred (by us) and later verified in *S.aureus*



[Antimicrob Agents Chemother.](#) 2003 Jun;47(6):2051-5.

FREE full text article at  
[aac.asm.org](#)

FREE full text article  
in PubMed Central

**Inhibitors of pantothenate kinase: novel antibiotics for staphylococcal infections.**

[Choudhry AE](#), [Mandichak TL](#), [Broskey JP](#), [Egolf RW](#), [Kinsland C](#), [Begley TP](#), [Seefeld MA](#),

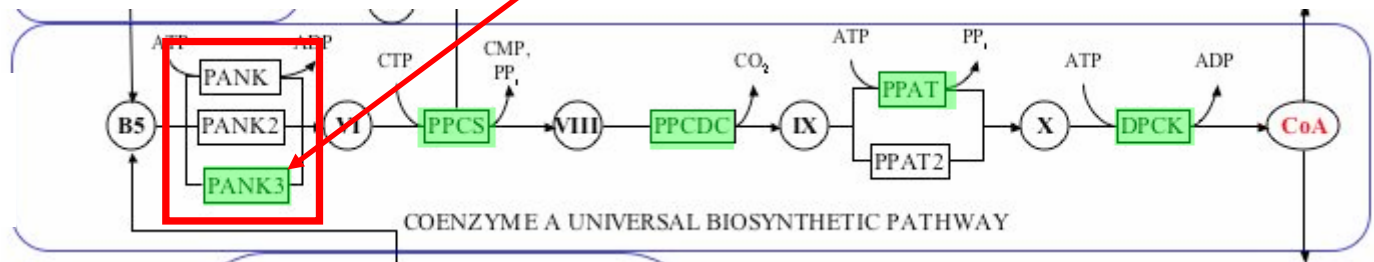


# Annotations of alternative pantotenate kinase genes in CoA\_Biosynthesis

7	PANK	Pantothenate kinase (EC 2.7.1.33)	<a href="#">R03018</a>
8	PANK2	Pantothenate kinase type II, eukaryotic (EC 2.7.1.33)	<a href="#">R03018</a>
9	PANK3	Pantothenate kinase type III, CoaX-like (EC 2.7.1.33)	<a href="#">R03018</a>
10	PPCS	Phosphopantothenoylcysteine synthetase (EC 6.3.2.5)	<a href="#">R04231</a> ,
11	PPCDC	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	<a href="#">R03269</a>
12	PPAT	Phosphopantetheine adenylyltransferase (EC 2.7.7.3)	<a href="#">R03035</a>
13	PPAT2	Phosphopantetheine adenylyltransferase, type II eukaryotic (EC 2.7.7.3)	<a href="#">R03035</a>
14	DPCK	Dephospho-CoA kinase (EC 2.7.1.24)	<a href="#">R00130</a>

Genome ID	Organism	Variant Code	*PANK	PPCS	PPCDC	*PPAT	DPCK	
83333.1 <a href="#">Color</a>	Escherichia coli K12 [B]	A	<a href="#">3890-7</a>	<a href="#">3575</a>	<a href="#">3575</a>	<a href="#">3570-12</a>	<a href="#">103</a>	model organisms
224308.1 <a href="#">Color</a>	Bacillus subtilis subsp. subtilis str. 168 [B]	B	<a href="#">2381-7</a> , <a href="#">70-9</a>	<a href="#">1572</a>	<a href="#">1572</a>	<a href="#">1504-12</a>	<a href="#">2909</a>	
158879.1 <a href="#">Color</a>	Staphylococcus aureus subsp. aureus N315 [B]	C	<a href="#">1999-8</a>	<a href="#">1089</a>	<a href="#">1089</a>	<a href="#">1004-12</a>	<a href="#">1561</a>	
261594.1 <a href="#">Color</a>	Bacillus anthracis str. 'Ames Ancestor' [B]	C	<a href="#">2974-8</a> , <a href="#">370-9</a>	<a href="#">2999</a> , <a href="#">3983</a>	<a href="#">3983</a>	<a href="#">4112-12</a>	<a href="#">4765</a>	Pathema organisms
149 <a href="#">Color</a>	Clostridium botulinum ATCC 3502 [B]	E	<a href="#">2254-9</a>	<a href="#">1240</a>	<a href="#">1240</a>	<a href="#">3223-12</a>	<a href="#">2054</a>	
272 <a href="#">Color</a>	Burkholderia pseudomallei K96243 [B]	E	<a href="#">2992-9</a>	<a href="#">3248</a>	<a href="#">3248</a>	<a href="#">2660-12</a>	<a href="#">5587</a>	

All of *Burkholderia* and *Clostridium* ssp rely solely on PANK3



Similar situation in *Clostridium botulinum* and *Burkholderia pseudomallei* K96243

NT02CB3552	NT02CB2499	NT02CB2480	NT02CB2480	NT02CB2999
?	ntbp0881	ntbp0881	ntbp0521	ntbp3008
transcriptional activator, putative, Baf family	phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine (ligase)	phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine (ligase)	phosphopantetheine adenylyltransferase	dephospho-CoA kinase

# Virulence related Subsystems in NMPDR

Proteins can be grouped based on various criteria:

Subsystems		
Classification	Subsystem	Curator
Virulence	<a href="#">Heme-bound Iron Scavenge Pathway</a>	master:RickS
	<a href="#">Listeria Pathogenicity Island LIPI-1 extended</a>	master:SvetaG
	<a href="#">Mannose-sensitive hemagglutinin type 4 pilus</a>	master:RobE
	<a href="#">Regulation of Oxidative Stress Response</a>	master:MikeK
	<a href="#">SLO-NADGH Locus</a>	
	genome context (virulence islands, prophages, conserved gene clusters)	
	virulence mechanism	
	<a href="#">Streptococcus pyogenes recombinatorial zone</a>	master:RamyA
	<a href="#">MLST</a>	master:RobE
	<a href="#">Streptococcal Hyaluronic Acid Capsule</a>	
	enzymatic activity	
	cellular localization	
	<a href="#">Motility of Campylobacter</a>	master:OlgaZ
	predicted or measured co-regulation	
	<a href="#">Streptococcus pyogenes virulence regulators</a>	
	common phenotype	
	<a href="#">Tetracycline resistance, ribosome protection type</a>	master:gjo_and_km
	<a href="#">Two-component regulatory systems in Campylobacter</a>	master:OlgaZ
	combinations of criteria	
	<a href="#">Streptolysin S Biosynthesis and Transport</a>	
	<a href="#">Cholera toxin</a>	master:VeronikaV

